Finally on Par?! Multimodal and Unimodal Interaction for Open Creative Design Tasks in Virtual Reality

Chris Zimmerer
University of Würzburg
Human-Computer Interaction
Würzburg, Germany
chris.zimmerer@uni-wuerzburg.de

Erik Wolf
University of Würzburg
Human-Computer Interaction
Würzburg, Germany
erik.wolf@uni-wuerzburg.de

Sara Wolf
University of Würzburg
Psychological Ergonomics
Würzburg, Germany
sara.wolf@uni-wuerzburg.de

Martin Fischbach
University of Würzburg
Human-Computer Interaction
Würzburg, Germany
martin.fischbach@uni-wuerzburg.de

Jean-Luc Lugrin
University of Würzburg
Human-Computer Interaction
Würzburg, Germany
jean-luc.lugrin@uni-wuerzburg.de

Marc Erich Latoschik
University of Würzburg
Human-Computer Interaction
Würzburg, Germany
marc.latoschik@uni-wuerzburg.de

Figure 1: Illustration of five products created with the MMI (first two) and the UMI (last three) in the open creative design task.

ABSTRACT

Multimodal Interfaces (MMIs) have been considered to provide promising interaction paradigms for Virtual Reality (VR) for some time. However, they are still far less common than unimodal interfaces (UMIs). This paper presents a summative user study comparing an MMI to a typical UMI for a design task in VR. We developed an application targeting creative 3D object manipulations, i.e., creating 3D objects and modifying typical object properties such as color or size. The associated open user task is based on the Torrence Tests of Creative Thinking. We compared a synergistic multimodal interface using speech-accompanied pointing/grabbing gestures with a more typical unimodal interface using a hierarchical radial menu to trigger actions on selected objects. Independent judges rated the creativity of the resulting products using the Consensual Assessment Technique. Additionally, we measured the creativity-promoting factors flow, usability, and presence. Our results show that the MMI performs on par with the UMI in all measurements despite its limited flexibility and reliability. These promising results demonstrate the technological maturity of MMIs and their potential to extend traditional interaction techniques in VR efficiently.

ICMI '20, October 25–29, 2020, Virtual event, Netherlands

ACM ISBN 978-1-4503-7581-8/20/10…$15.00
https://doi.org/10.1145/3382507.3418850

CCS CONCEPTS

• **Human-centered computing → Virtual reality**; **Interaction techniques**; *Empirical studies in HCI*.

KEYWORDS

Multimodal Interaction; Speech and Gesture; 3D User Interfaces; Virtual Reality; Creativity; Design; User Study

ACM Reference Format:
Chris Zimmerer, Erik Wolf, Sara Wolf, Martin Fischbach, Jean-Luc Lugrin, and Marc Erich Latoschik. 2020. Finally on Par?! Multimodal and Unimodal Interaction for Open Creative Design Tasks in Virtual Reality. In *Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20), October 25–29, 2020, Virtual event, Netherlands.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3382507.3418850

1 INTRODUCTION

Bolt pioneered instruction-based Multimodal Interfaces (MMIs) for large graphical displays [8]. Such interfaces promoting a synergistic use of speech and gestures have also been considered to provide promising interaction paradigms for Virtual Reality (VR) [32, 40, 42, 43, 49]. The potential benefits of MMIs include increased expressiveness, flexibility, reliability, and efficiency [61, 64, 65, 75]. Nevertheless, MMIs are still used considerably less often than Unimodal Interfaces (UMIs) in VR. Typical 3D user interfaces for system control tasks consist of graphical menus and spatial 3D input devices such as physical controllers that combine push-buttons and joysticks with 3D position and rotation tracking [48].

Possible explanations for this lopsidedness are the technological challenges of MMIs regarding the recognition of probabilistic user input [32, 88], their integration in real-time interactive systems [21, 46], and semantic integration [44, 45, 66]. The technological maturity announced for MMIs ten years ago [39] applies to multi-touch devices, but does not extend to synergistic speech and gesture interfaces for VR [65, pp. 449–478]. However, considerable progress has been made in recent years: (1) Commercially available, low-cost hardware like the Oculus Rift [77] head-mounted display and software solutions like the Unity [78] game engine help to make VR more accessible for the research community and the general public. (2) Machine learning has considerably improved the unimodal recognition of speech and gesture through large data sets and deep learning approaches [12, 54]. As a result, the recognition rate of multimodal systems increased as well. (3) Newly developed software concepts facilitate the implementation of MMIs specifically for real-time interactive systems [21, 88]. In summary, all these accomplishments advanced the technological maturity of MMIs for VR [85, 90], but an evaluation of the progress made is still pending.

The technological maturity is only one aspect that promotes the use of MMIs. It is equally important to understand how the technological means can be used to develop effective and efficient MMIs. There is a large body of research and existing guidelines focusing on UMIs (see LaViola Jr et al. [48] for an overview). However, guidelines for MMIs are still sparse. In particular, it is not yet clear which combinations of modalities best accomplish certain tasks in specific application areas [63, 65]. This research gap stems in part from a lack of studies that compare unimodal and fully implemented multimodal interfaces in different domains. This lack is in turn associated with the technological challenges of implementing MMIs [20]. One notable exception is the work of Oviatt, who showed that a multimodal —speech and pen— interface is more efficient than an unimodal solution in a dynamic interactive map system [58, 65]. Altogether, it remains an open question whether fully implemented MMIs can finally perform on par with, or even better than, UMIs, given the technological progress and their potential benefits.

In the present work, we chose design tasks as the target application domain. Such tasks especially benefit from VR in terms of efficiency and effectiveness through lower costs, improved configuration options, dynamic simulations, and possibilities for collaboration [17, 25, 83]. For example, the VR building platform IrisVR [29] helped to detect and correct design inconsistencies in the construction of water recycling centers [28]. However, there is almost no research investigating the influence of the user interface on both the creative process and the creative product, despite its theoretical importance [36]. In our previous work [85], we showed that a multimodal –speech and gesture– interface outperforms a typical unimodal –menu-based– interface in a VR object modification task with regard to the creativity-promoting characteristics flow, usability, and presence. However, our evaluation task at the time did not actually require the participants to be creative themselves. In our present work, we use a creativity demanding design task and provide the missing evaluation of the interfaces' influence on the creative product. With a partial reproduction of our previous work, we showcase one aspect of technical maturity, enhance comparability in the discussion of the results, and strengthen a necessary basis for deriving applicable design guidelines for MMIs in the future.

Contribution: We present a summative user study comparing a fully implemented multimodal –speech and gesture– interface with a typical unimodal –menu-based– interface in an open creative VR design task. The task is based on the Torrence Tests of Creative Thinking [81]. Our results show that the MMI is rated as good as the the UMI regarding the creativity-promoting factors flow, usability, and presence. In addition, the Consensual Assessment Technique [1] did not reveal any differences between products created with the MMI or UMI in terms of their judged creativity. The MMI performs on par with the UMI despite its significantly lower reliability (i.e. more recognition errors) and limited flexibility (i.e. constrained grammar and limited vocabulary). Altogether, our results demonstrate not only the technological maturity of MMIs, but also contribute towards the establishment of design concepts and guidelines for MMIs. We provide a detailed evaluation of a fully implemented MMI in VR from which we derive a concrete design recommendation and propose two generic guidelines for future research and development.

2 RELATED WORK

In the following, we give an overview of the current state of 3D user interfaces for VR and subsequently highlight established techniques for measuring creativity.

2.1 User Interfaces for VR

Integrated 2D graphical menus operated by unimodal physical input devices are typical 3D user interfaces in VR for performing system control tasks [15, 48]. For example, users instruct the system to modify a 3D object by choosing an action from a graphical menu with a joystick and the press of a button [29]. The large number of freely available plugins for unimodal interaction techniques in VR shows not only their technological maturity but also the comprehensive research that went into their design. For instance, the XR Interaction Toolkit [79] or the Virtual Reality Toolkit [80] for the Unity game engine contain a wide variety of interaction techniques, including controller input and different types of menus.

Radial menus are especially suitable regarding efficiency, usability, and error rates [11, 16, 22, 72]. They are usually composed of a disk separated into equal segments where each segment represents a system command. More complex implementations support the hierarchical organization of menu entries. Choosing a menu entry triggers a sub-menu for specifying further parameters to complete a system command [22]. However menus are not without drawbacks. They force users to shift their visual attention between the objects of interest and the menu [69], potentially breaking the users' flow. Depending on the display type VR may cause a vergence-accommodation conflict that amplifies this negative effect [38]. Additionally, even the efficiency of radial menus decreases with increasing complexity and number of menu-items [16].

In contrast to unimodal interactions, multimodal interactions commonly combine at least two modalities potentially operating simultaneously [57]. The specification of parameters required for triggering system actions can thus be distributed among adequate modalities synergistically [32], e.g., speech and gesture. Users can stay focused on the objects of interest [62], which decreases the need for attention shifts. MMIs show advantages in usability, such as increased efficiency [48, 70] and user satisfaction [18, 59]. Further, MMIs potentially induce less cognitive load than UMIs [64].

Figure 2: The participants were asked to design an object that represents the feeling of joy. They could create, delete, move, and modify simple 3D objects. Creation, deletion, and modifications of objects' size and color were performed with either a unimodal (radial menu) or multimodal (speech and gesture) interface. The two images on the left show the two-step process of changing the color of an object using the UMI. The image on the right depicts the same action using the MMI.

However, only a few studies compare unimodal and fully implemented multimodal techniques and measure their effects on activities that require high cognitive resources such as creative performance [36]. In a previous experiment [85], we confirmed the expected advantages of multimodal interaction in VR for the creativity-promoting factors usability, flow, and presence in a standardized object modification task. Thus, our current work continues this line of research by using a creativity-demanding VR design task and providing measurements regarding the designed products.

2.2 Measuring Creativity

Evaluating creativity empirically remains problematic due to the criterion problem, which is a direct result of the field's complexity and multidimensionality [68]. Creativity has not yet been described in its entirety by one grand theory [6]. A potential theory, including every aspect of creativity, is even deemed so cumbersome that it would be incomprehensible and of no use in practical research [35]. However, there are two recurring aspects in most definitions of creativity: novelty and usefulness, or sometimes also called originality and effectiveness/appropriateness [24, 50, 71, 74]. Thus, a product, object, or idea is creative when it is judged as novel and useful. In turn, creative people are people who are capable of creating novel and useful products.

Amabile [1] introduced the Consensual Assessment Technique (CAT), to compensate for the lack of an operational definition. This technique is based on a consensual definition of creativity regarding the product and not regarding the process or person [68]. It states that a product is creative to the extent that judges independently agree on it. Thus, the technique's validity is not linked to a particular theory or definition of creativity, but to the subjective definition of each judge and the inter-rater reliability between all judges. The product's creativity is measured by the extent to which judges agree that one product is more creative than another. The CAT overcomes the difficulty of defining objective criteria for identifying creativity in products. It solely relies on the subjective criteria of appropriate judges. CAT is well-validated and widely used in creativity research [2, 3, 7, 68]. When applying CAT, important factors regarding the (1) judges and the (2) task have to be considered. (1) The judges must be sufficiently familiar with the product domain in order to have developed some implicit subjective criteria for creativity in this domain [4]. However, the degree of familiarity does not have to be the same for all judges [2]. In specialized domains, expert judges cannot be easily replaced by non-experts. For instance, in the domain of poetry, college students showed less inter-rater reliability than professional poets when judging poems [34]. Creativity in less specialized domains like paper collages can be rated by non-experts from various backgrounds [2]. (2) The task should be open-ended enough to permit flexibility in outcomes and the creation of novel products allowing the user to be creative [1]. It shall not depend on specialized skills, e.g., ability to draw, to avoid large individual differences in baseline performance.

The Torrence Tests of Creative Thinking (TTCT) [81] propose widely used tasks that satisfy the aforementioned requirements of CAT [1]. They are divided in verbal and figurative tasks, which aim to elicit the participants' creativity. For instance, in a figural task, the objective is to create a paper collage by combining simple objects of various sizes and colors using scissors and glue. This task is usually performed on paper and, to the best of our knowledge, has not yet been performed in a fully-immersive HMD-based VR application. The TTCT are a suitable tool for measuring creativity in gifted people and in the general population [37].

While CAT evaluates creativity by focusing on the creative product, there are other concepts associated with the creative process in VR. In particular, people engaged in creative tasks often reported a feeling of flow [86]. Flow is achieved if the users can fully invest their attention in the task at hand and if the task's difficulty matches their abilities [13]. Accordingly, the interface must be as usable as possible to minimize disruptions caused by its utilization and to introduce as little additional workload as possible [67]. This can be achieved by leveraging users' experience, knowledge, and engrained behavioral patterns when designing the interface [60]. Intuitive use, a sub-concept of usability, is of particular importance in this context. It suggests that interfaces are more effective, satisfying, and require fewer cognitive resources if a user can operate them by subconsciously applying previously acquired knowledge [26, 56]. For example, head-tracking in VR is considered to be very intuitive, since manipulating one's viewpoint by moving the head is a natural and well-known interaction from the real world. Thus, high usability may foster flow and subsequently creativity. Jin provides empirical evidence that presence plays a mediating role in inducing flow [30, 31]. For this reason, we also consider it to be an important factor for creativity. Taken together, flow, usability, and presence can influence creative processes in VR. Interaction techniques should promote these factors in VR design applications.

Table 1: **Overview of our variables and hypotheses categorized in creative process and creative product.**

Variable	Hypotheses
Creative Process	
(H1) Flow	UMI < MMI
(H2) Usability	UMI < MMI
(H3) Presence	UMI < MMI
Creative Product	
(H4) Creativity Assessment	UMI < MMI

3 STUDY

3.1 Approach and Hypotheses

To the best of our knowledge, there is no prior research regarding the assessment of creativity in creativity-demanding VR design tasks. Therefore, we propose an approach based on the Torrance Tests for Creative Thinking [81] and the Consensual Assessment Technique [1]. We created an open VR design task based on one of the figural tasks of the TTCT. Participants were asked to create a three-dimensional object representing something "joyful" in a virtual environment by creating, modifying, and combining primitive forms, i.e., spheres, boxes, and pyramids (see Figure 2). We chose the concept "joy" since it represents a commonly understood feeling and has been used successfully in previous research [1]. The available actions ranged from creating and deleting objects to changing their position, orientation, size, and color. Thus, participants were able to create novel products without any specialized skills in an open-ended, flexible, building-block-based design task.

The study was conducted in a between-subject design where participants either used the MMI or the UMI. We evaluated both the creative process by measuring the creativity-promoting factors flow, usability, and presence, and the creative products by using the CAT. Based on the results of our previous experiment [85], we expected that our implementation of the MMI supports a higher feeling of flow, usability, and presence compared to the UMI. Subsequently, we assumed that the products created with the MMI are judged as more creative. The hypotheses are presented in Table 1.

3.2 Apparatus

3.2.1 Virtual Environment. We implemented a virtual environment, a UMI, and an MMI to test our hypotheses. As a basis, we reused the implementation of our previous work [85]. The virtual environment was realized with Unity 2017.4.8f1 [78]. We used an Oculus Rift S HMD [77] for visualization, two Oculus Touch Controllers, and the built-in microphone for interaction. Both interfaces relied only on this consumer-level hardware to increase comparability.

The environment consisted of a simple room enabling realistic object shape, color, and depth perception while not distracting participants from the experimental task. It featured a centered one-meter high podium on which participants created their products (see Figure 1). Object gravity was disabled, and objects could arbitrarily intersect with each other. Virtual hands and controllers were displayed in VR (see Figure 2). The application ran on a VR-capable

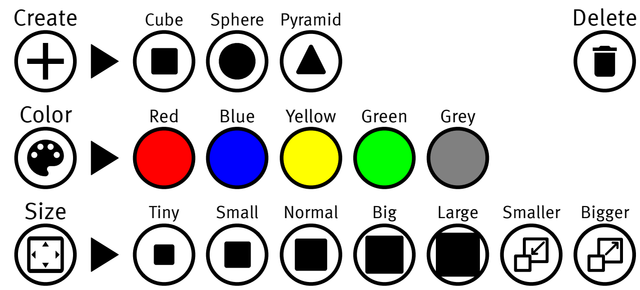

Figure 3: A depiction of the icons used in the radial menu. The left side shows all possible actions, the right their respective parameters. The delete action in the upper right corner has additional parameters.

PC that allowed for fluent rendering. We logged the framerates and registered no noticeable drop below 90 fps during the experiment.

3.2.2 General Interface. In both the MMI and UMI condition, participants used the same interactions for selection and object movement. The ray-casting and virtual-hand technique was used for selection [48]. A virtual object was selected as long as the user's virtual hand or the ray (cast from the virtual hand into the environment) intersected with it. A selection was signaled with soft controller vibrations and a white frame around the selected object. Only the virtual-hand technique was used for object movement. First, the user had to place their virtual hand inside the object. Second, pressing and holding the trigger button attached the object to the virtual hand. Third, the user was able to drag the object to a new position by moving the virtual hand. Fourth, releasing the trigger button detached the object from the user's virtual hand.

The MMI and UMI were used for object creation, modification, and deletion. When creating an object, it spawned at the position of the dominant hand. An object had to be selected beforehand to perform an object modification or deletion. Participants could always change the last modified object without selecting it again. In both the UMI and MMI, a help menu (triggered with the B-button) provided information about how to use the respective interface.

3.2.3 Unimodal Interface. The UMI consisted of a two-level radial menu with custom icons and was implemented using the Virtual Reality Toolkit V3.2.0 [84] (see Figure 2, left). It was bound to the non-dominant virtual hand and could be opened by pressing the controller's joystick. The first level of the menu provided all possible actions. The second level opened after selecting an action by pressing the joystick. It displayed corresponding parameters, e.g., the colors after choosing the color action. All actions and their parameters are summarized in Figure 3. After selecting a parameter, the menu closed automatically and the system performed the action. Deletion was the only action directly executed for the currently selected object with no further parameters. Menu icons were highlighted by moving the controller's joystick and selected by pressing it. The hand trigger served as the back and close button.

3.2.4 Multimodal Interface. The MMI consisted of a synergistic combination of speech and pointing/grabbing gestures. The interface was implemented using the open-source platform Simulator X

Table 2: An overview of all multimodal commands using speech and pointing/grabbing gestures. All commands are translated to English but were used in German in the study.

Create	Utterances	Create a <Property>.
	Property	cube | sphere | pyramid
Size	Utterances	Make <Object> <Property>. | Make <Object> smaller/bigger.
	Object	that *[pointing, grabbing]* object | that *[pointing, grabbing]* | it
	Property	tiny | small | normal | big | large
Color	Utterances	Paint <Object> <Property>.
	Object	that *[pointing, grabbing]* object | that *[pointing, grabbing]* | it
	Property	red | blue | yellow | green | gray
Delete	Utterances	Delete <Object>.
	Object	that *[pointing, grabbing]* object | that *[pointing, grabbing]* | it

[21, 47]. Simulator X contains an implementation of a concurrent Augmented Transition Network (cATN) [88] and is freely available [89]. The cATN is the successor of the temporal Augmented Transition Network (tATN) [41] and was used to define and recognize possible multimodal utterances. We used the Microsoft Speech SDK for speech recognition [52]. An in-depth description of the system architecture can be found in our previous work [85]. All supported utterances are displayed in Table 2. For instance, changing the color of an object required the following interaction: First, the user had to utter the corresponding keyword "paint" to denote the type of action. Second, an object had to be selected by either pointing or grasping while simultaneously saying "that object" or simply "that". Last, the new color was defined by speech, e.g., "red". A pulsing microphone icon in the participant's field of view provided feedback on active speech and gesture processing.

3.3 Measurements

Our measurements are categorized into dependent variables regarding the creative process and the creative product. The former consists of measurements regarding flow, usability, and presence. The latter consists of the judges' evaluation of the designed products in terms of creativity, novelty, complexity, and effort. We captured video, sound, and screen recordings of all sessions. We logged the time each participant spent designing their product and all performed actions. In addition, we recorded control variables to reveal and control differences between both conditions.

3.3.1 Creative Process. To measure flow, we used the FQNR, a survey on the two flow characteristics enjoyment and concentration [23]. It captures four items, each on a Likert scale ranging from 1 to 5 (*5 = high feeling of flow*). In addition, we employed the GEQ subscale for flow [27]. It consists of five items, each on a Likert scale ranging from 0 to 4 (*4 = high feeling of flow*). Lastly, we recorded the relative subjective duration (RSD) [14], which was calculated using Equation 1.

$$\text{RSD} = \frac{\text{Perceived task duration} - \text{Actual task duration}}{\text{Actual task duration}} \quad (1)$$

The lower the calculated percentage, the higher the feeling of flow. RSD builds upon the assumption that a high level of flow provides the impression that time passes faster.

Regarding usability, we measured intuitive use, workload, efficiency, and the effectivity of the MMI. We used the Questionnaire for Subjective Consequences of Intuitive Use (QUESI) [55]. The QUESI captures 14 items, each on a Likert scale ranging from 1 to 5 structured in five dimensions: mental workload, the achievement of goals, the perceived effort of learning, familiarity, and the perceived error rate. The total score ranges from 1 to 5 (*5 = high intuitiveness*). The mental workload was measured with the SEA scale [19], the german version of the Rating Scale Mental Effort [5, 87]. It consists of a single item ranging from 0 to 220 (*220 = high workload*) presented after the experimental task was finished. To capture efficiency, we calculated the actions per minute (APM) from the total number of performed actions and the total task duration. In addition, we counted the recognition errors of the MMI post-hoc from the recordings. We logged the following errors: The MMI recognized a wrong action, a wrong parameter, or both (e.g., "paint it red" instead of "make it small"), it recognized no valid command when it should have, and it recognized a valid command when it should not have. These errors were only logged for the MMI, since they do not occur in a non-probabilistic UMI. We calculated the errors per action (EPA) based on the counted recognition errors and the total number of actions performed to obtain an indicator for the interfaces' effectivity.

Lastly, we used the PQNR introduced by Bouchard et al. [9, 10] to measure presence. It consists of a single item scale ranging from 0 to 10 (*10 = high presence*).

3.3.2 Creative Product. The designed products were evaluated by judges following the CAT. Creativity was measured on more than one dimension, as recommended by Amabile [2]. Judges were asked to evaluate each product regarding creativity, novelty, complexity, and the effort put into its creation. Each dimension was rated using a single item on a 5-point Likert scale ranging from 1 to 5 (*5 = very creative, novel, etc.*). The questions were taken from Amabile [3].

3.3.3 Control Variable. We used the Kaufman Domains of Creativity Scale (K-DOCS) [33, 51] to assess the participants' creativity. The K-DOCS measures self-rated creativity on a five-point Likert scale (*1 = much less creative*; *5 = much more creative*). Participants were asked to state their creativity in comparison to people of their age and life experience using 50 behavior-based questions that reflect a domain-specific perspective of everyday creativity. Since we did not expected any simulator sickness [73], we measured the participants' well being with a single question: *Do you currently feel physical discomfort?*.

3.4 Participants

3.4.1 Creative Process. For the interface comparison, 56 participants were randomly assigned to either the UMI or the MMI condition. Participants received one hour of course credit or financial compensation. Participants (16 male, 40 female) were aged between 19 and 62 years (M = 27.29, SD = 9.76). They had no hearing impairments and normal, or to normal corrected, vision. All participants were native speakers or spoke the German language for more than

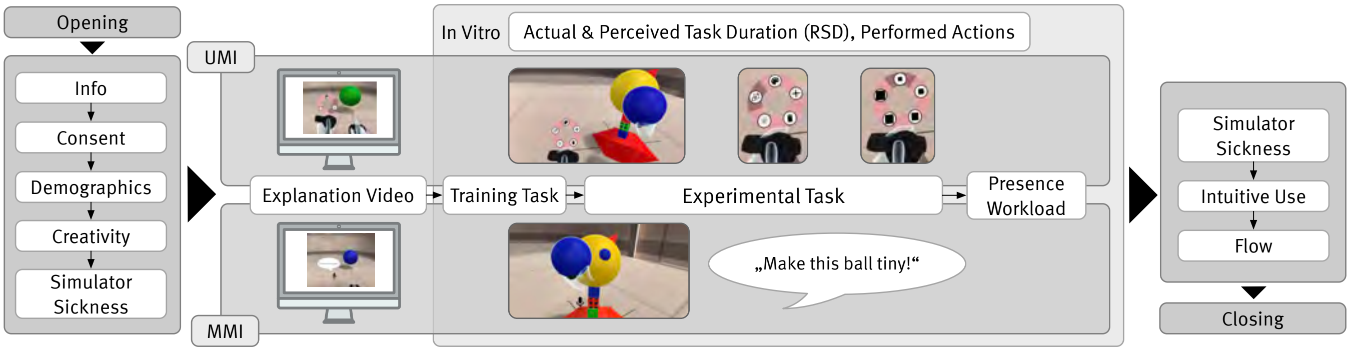

Figure 4: An overview over the experimental procedure of the interface comparison which is read from left to right and top to bottom. A detailed description of the procedure can be found in Section 3.5.1.

ten years, which was deemed sufficient to use the MMI. Five participants experienced VR for the first time, 42 experienced VR one to ten times, and nine experienced VR more than ten times. In the UMI condition, we tested 28 participants aged between 19 and 62 (M = 27.86, SD = 11.31), and in the MMI condition, we tested 28 participants aged between 19 and 56 (M = 26.71, SD = 8.08). To control the distribution of our tested samples, we compared both conditions regarding age, gender, and creativity (K-DOCS). Pairwise comparisons between all dimensions of these factors did not show significant differences.

3.4.2 Creative Product. 18 judges (5 male, 13 female) between the age of 19 and 22 (M = 20.22, SD = 1.17) rated the designed products. All judges were students at a university, native speakers, and all but one experienced VR in the past.

3.5 Procedure

3.5.1 Creative Process. Participants followed a strict experimental procedure depicted in Figure 4. In the beginning, they read the experimental information, gave consent, and generated a code for pseudonymization. Then, they answered questions on demography, creativity (K-DOCS), and simulator sickness. Participants got familiar with the respective interaction technique by watching an explanation video and performing a five-minute training phase. During the training in VR, participants were asked to replicate a simple object which required each possible interaction to be performed at least once. The test phase consisted of the prior explained task instructed as follows: “Please create an object that represents the feeling of joy”. Participants decided by themselves when their object was finished. We imposed no time limits for the experiment. During the training and test phase, we logged framerate, task duration, number of actions, and what kind of actions were performed and started the video, audio, and screen recording. Upon finishing the task, we captured the mental workload and presence scores while participants were still in VR. In addition, we asked participants to indicate the perceived task duration before they had the opportunity to check the time. Combined with the measured task duration, we calculated the RSD. All these instructions and questions were audio-recorded and played in VR. We instructed participants to answer the questions verbally while the experimenter logged each answer. After leaving the VR, participants answered

Table 3: This table shows the absolute number of performed actions for both interfaces.

	Performed Actions				
	Create	Delete	Scale	Color	**Total**
UMI	682	58	745	692	2177
MMI	691	63	763	654	2171

the post-question on simulator sickness and the questionnaires on intuitive use and flow. Depending on the task duration, the whole experimental procedure took between 40 and 60 minutes.

3.5.2 Creative Product. The evaluation was carried out by using an online survey that was structured to meet the requirements of the CAT procedure [2, 4]. First, the judges had to fill in a demographic questionnaire. Second, they were presented with instructions on how the products were created and how they should be evaluated based on their subjective criteria of creativity. We did not provide the judges with any definition of creativity to not influence them. Further, they were instructed to rate the products relative to each other and not against an absolute standard, e.g., a famous sculpture of a professional artist. Third, judges viewed short videos (15 seconds) of each product without rating them to get a holistic impression of all products. Lastly, they viewed every video again and rated the product on the four dimensions: creativity, novelty, complexity, and effort. The order in which the videos were presented to the judges was assigned randomly to avoid order effects.

4 RESULTS

Table 3 depicts the distribution of the absolute number of actions performed per interface. Participants using the UMI performed on average (M = 77.75, SD = 41.54) a similar amount of actions as participants using the MMI (M = 77.54, SD = 37.84), $t(25) = 0.02, p = .984$. Similarly, the average time for object completion in the UMI condition (M = 10.89 minutes, SD = 5.37) was not significantly different than in the MMI condition (M = 10.93 minutes, SD = 5.44), $t(25) = 0.03, p = .978$.

Table 4: The table shows the results of the statistical analysis regarding the differences in flow, usability, and presence as well as the for the creativity assessment measurements between both interfaces.

	UMI (n = 28)	**MMI** (n = 28)	
	M (SD)	M (SD)	p
Flow (H1)			
FQNR score	4.36 (0.44)	4.38 (0.51)	.435
GEQ score	2.84 (0.58)	2.77 (0.49)	.308
RSD in %	13.34 (35.52)	6.64 (38.78)	.252
Usability (H2)			
QUESI score	3.51 (0.71)	3.54 (0.51)	.252
SEA score	53.46 (28.82)	55 (35.64)	.409
APM	7.40 (2.48)	7.37 (1.79)	.347
EPA	0.0 (0.0)	0.15 (0.08)	$< .001^{*}$
Presence (H3)			
PQNR score	8.00 (1.09)	8.19 (1.33)	.197
Creativity Assessment (H4)			
Creativity	3.23 (0.73)	3.14 (0.71)	.653
Novelty	2.97 (0.72)	3.00 (0.77)	.897
Complexity	2.80 (0.62)	2.73 (0.57)	.692
Effort	2.83 (0.71)	2.83 (0.68)	.966

* indicates significant results

For our statistical analysis, we corrected values below the 5th and above the 95th percentile by the use of the Winzorizing correction approach [82] to deal with the few outliers in the data (0.03 % of all values). After adjusting outliers, all data showed homogeneity of variances in the performed Levene's tests. However, Shapiro-Wilk tests revealed a violation of normal distribution for the data of the FQNR score, APM, SEA score, and PQNR score. Therefore, we calculated two-sided Mann-Whitney-U tests for the measurements with violated pre-assumptions and two-sided t-tests for all other measurements. For CAT measurements, we calculated consistency and reliability among the judges using Cronbach's α [76]. Judges' ratings were found to be highly consistent for creativity (α = .904), novelty (α = .907), complexity (α = .868), and effort (α = .919). Table 4 summarizes the descriptive data and test results for the interface comparison. An exploratory analysis of the relationship between EPA and flow, usability, and presence showed no signs of an influence of recognition errors on the measurements. Further exploratory analysis showed no gender differences. The simulator sickness question showed no signs of discomfort in a pre-post comparison.

Additionally, we explored the assumption that creative performance is related to flow, usability, and presence. To this end, we calculated the correlations between the variables of the creativity rating and the variables of the aforementioned factors for the entire sample (n = 56). Since the data for FQNR score, APM, SEA score, and PQNR score violated normal distribution, we calculated Spearman's ρ for these variables. For the remaining variables, we calculated Pearson's r. Table 5 shows the correlation matrix.

Table 5: The table shows the correlation matrix between flow, usability, and presence variables and the creativity assessment variables.

	Creativity Assessment			
	Creativity	Novelty	Complexity	Effort
Flow				
FQNR score	$.28^{*}$	.22	$.34^{*}$	$.35^{\dagger}$
GEQ score	.14	.02	0.22	$.28^{*}$
RSD in %	.11	.08	.08	.09
Usability				
QUESI score	$.31^{*}$	.19	$.28^{*}$	$.32^{*}$
SEA score	$-.35^{\dagger}$	$-.38^{\dagger}$	$-.27^{*}$	$-.17$
APM	.14	.08	.12	.18
EPA	$-.07$	$-.01$	$-.01$	.20
Presence				
PQNR score	.18	.10	.10	.09

$^{*}p < .05$, $^{\dagger}p < .01$

4.1 Findings Summary

Surprisingly, the most important finding is the equivalent performance between the interfaces, despite the significantly lower reliability as well as limited flexibility of the MMI. Indeed, we hypothesized that participants using the MMI perceive a higher feeling of flow (H1), a higher usability (H2), and a higher feeling of presence (H3) during the open creative task. As these factors are supposed to promote creativity, we hypothesized that products designed with the MMI are rated as more creative by independent judges (H4). Contrary to our hypothesis, our primary results show no significant differences regarding flow (H1), usability (H2), presence (H3), and the judged creativity of designed products (H4). The MMI performed on par with the UMI in all measurements

Another important finding are the significant correlations between the creativity assessment of designed products and the flow & usability measurements with moderate correlation coefficients. These correlations support the current state of research that the perceived feeling of flow and the usability of the utilized interface influence the creativity of the designed product. Ultimately, they emphasize the importance of highly usable interfaces to maximize creativity in VR design applications and thus the potential relevance of MMIs in this domain. The implications of these findings are discussed in the following sections.

5 DISCUSSIONS

We identify two reasons why the MMI has not achieved its hypothesized superiority over the UMI: (1) its low reliability (i.e., high recognition errors) and (2) its limited flexibility (i.e., constrained grammar and limited vocabulary). For (1) the MMI failed to correctly recognize the command every seventh interaction, while the UMI's button presses produced, as expected, no errors. An initial informal analysis of our system revealed that almost all recognition errors are due to the speech recognition error rate of the aforementioned off-the-shelf products. Despite these errors, the MMI was rated as good as the UMI. For (2), the MMI has been designed with a simple grammar (i.e. command structure) and vocabulary. It did not allow synonyms or alternative sentence structures. This substantially limited two commonly advocated advantages of MMIs: flexibility and naturalness.

However, further differences can explain the contradictory results from our last experiment [85]. The two main differences are the task the participants had to perform in the experiment and the study design. In the current experiment participants had to employ their creativity to create new objects from scratch in an open design task, instead of replicating presented objects. This has led to completely different results regarding flow, usability, and presence. Participants in the current experiment reported an overall higher feeling of flow (with both interfaces) than the ones in the previous experiment. This was to be expected since intrinsically motivated and demanding creativity tasks achieve a higher feeling of flow compared to extrinsically motivated monotonous replication tasks [53]. However, a higher feeling of flow may obscure subtle differences in otherwise very usable interfaces (cf. high QUESI scores) since participants are less focused on the interface and more on applying their creativity. In addition, the two study designs also differed. We used a *between-subject* design in contrast to the previously used *within-subject* design. Participants could not directly compare and judge between interfaces which could have had an additional influence on the their assessments.

All these differences raise interesting questions about the reliability of such comparisons and the conclusions drawn from them. The advantages or disadvantages of one interaction style over the other are heavily dependent on boundary conditions, e.g., task, environment, concrete implementation details, and dependent variables. Subtle differences in these boundary conditions seem to make the overall setups either too sensitive or too agnostic to changes in the independent variables.

6 LIMITATIONS AND FUTURE WORK

Our results show that the UMI and the MMI are suitable for the use in creative design applications. Nevertheless, the present work has several limitations that leave space for future work. First, the panel of judges in the CAT consisted only of students. Judges showed a high degree of agreement in their ratings, which is commonly regarded as the criterion for validity of the CAT. However, a comparison to a panel of judges consisting of professional artists, ideally with ample VR experience, could yield additional insights.

Second, our application supported comparatively fewer actions than actual design applications. We chose this setup since it was the first study using an open design task in VR. An in-depth investigation is necessary to determine whether a larger number of available actions negatively affects the performance of the MMI, as is the case with the performance of the radial menus in UMIs [16].

Lastly, as previously discussed, future work has to further research the generalizability of our results on other tasks, domains, and setups towards the goal of creating more generally applicable guidelines. In particular, comparisons should feature more complex MMIs that provide flexible and reliable natural interactions. However, due to the aforementioned lack of guidelines, such interfaces can not yet be straightforwardly implemented.

7 IMPLICATIONS AND GUIDELINES

Our results cast new light on the usability and applicability of multimodal interfaces in VR. Overall our findings and discussions can be formulated as one specific recommendation and two generic guidelines for future research and development:

Recommendation: **Consider a simple synergistic multimodal interfaces using speech-accompanied pointing & grabbing gestures in VR instead of hierarchical radial menus for 3D object manipulation in creative tasks**.

Guidelines: **Consider a simple synergistic multimodal interface for object modeling in VR**: Indeed, these types of interfaces are possible in VR and they are usable in VR, even when providing a limited flexibility and reliability. Both our current and previous work [85] demonstrate the technological maturity of MMIs in VR and address the current lack of comparative studies. We implemented a fully functional MMI with current software and hardware. Advanced hardware, better unimodal recognition systems, and new software concepts have not yet closed the technological gap between these interfaces, but have considerably narrowed it. We demonstrated that MMIs became technologically feasible for widespread use in VR applications and that recognition errors and limited flexibility do not render them completely unusable and are tolerated by the user.

Guidelines: **Consider summative studies between unimodal and multimodal interfaces**. There is still a research gap with regard to the development of truly flexible and natural MMIs for VR. This gap can be attributed to a lack of applicable guidelines which specify what modalities (and their combined, potentially synergistic use) are most likely to be beneficial for particular tasks in various application areas. Overall, the community needs more comparisons with different boundary conditions to build an extensive body of research from which more concise design guidelines for MMIs can be derived. Especially since their technological maturity makes it easier to conduct these comparisons. We especially emphasize the need for more comparative studies, which must carefully consider all boundary conditions like the task, design, and procedure, in order to be able to interpret and generalize results more thoroughly.

8 CONCLUSION

In this paper, we present a summative user study comparing a synergistic multimodal –speech and gesture– interface with a typical unimodal –menu-based– interface in an open creative VR design task. The MMI has been fully implemented with openly-available soft- and consumer hardware. We adapted a creativity demanding design task from the Torrence Tests for Creative Thinking to VR and used the *Consensual Assessment Technique* to rate the creativity of designed products. The MMI and the UMI achieved comparable and overall good scores in all measurements. The MMI induced comparable flow and presence and was rated as usable as the UMI despite its limited flexibility (i.e. constrained grammar and limited vocabulary) and its lower reliability (i.e. significantly higher recognition errors rate). The results demonstrate the technological maturity of MMIs in VR as well as their potential to extend traditional interaction techniques. Our future work will focus on developing and comparing more sophisticated MMIs and UMIs for different types of VR applications.

REFERENCES

[1] Teresa M Amabile. 1982. Social psychology of creativity: A consensual assessment technique. *Journal of personality and social psychology* 43, 5 (1982), 997.

[2] Teresa M Amabile. 1983. The social psychology of creativity: A componential conceptualization. *Journal of personality and social psychology* 45, 2 (1983), 357.

[3] Teresa M Amabile. 2018. *Creativity in context: Update to the social psychology of creativity.* Routledge.

[4] Teresa M Amabile and BA Hennessey. 2011. Consensual assessment. In *Encyclopedia of Creativity, Second Edition*, Runco MA and Pritzker SR (Eds.). San Diego: Academic Press, vol.1, pp. 253–260.

[5] Albert G. Arnold. 1999. Mental Effort and Evaluation of User-Interfaces: A Questionnaire Approach. In *Proceedings of HCI International (the 8th International Conference on Human-Computer Interaction) on Human-Computer Interaction: Ergonomics and User Interfaces-Volume I - Volume I.* L. Erlbaum Associates Inc., USA, 1003–1007.

[6] John Baer. 2011. Why grand theories of creativity distort, distract, and disappoint. *The International Journal of Creativity & Problem Solving* (2011).

[7] John Baer and Sharon S McKool. 2009. Assessing creativity using the consensual assessment technique. In *Handbook of research on assessment technologies, methods, and applications in higher education.* IGI Global, 65–77.

[8] Richard A. Bolt. 1980. "Put-That-There": Voice and Gesture at the Graphics Interface. *SIGGRAPH Comput. Graph.* 14, 3 (July 1980), 262–270. https://doi.org/10.1145/965105.807503

[9] Stéphane Bouchard, Geneviéve Robillard, Julie St-Jacques, Stéphanie Dumoulin, Marie-Josée Patry, and Patrice Renaud. 2004. Reliability and validity of a single-item measure of presence in VR. In *Proceedings of the Second International Conference on Creating, Connecting and Collaborating through Computing.* IEEE, 59–61.

[10] Stéphane Bouchard, Julie St-Jacques, Geneviève Robillard, and Patrice Renaud. 2008. Anxiety increases the feeling of presence in virtual reality. *Presence: Teleoperators and Virtual Environments* 17, 4 (2008), 376–391.

[11] Jack Callahan, Don Hopkins, Mark Weiser, and Ben Shneiderman. 1988. An empirical comparison of pie vs. linear menus. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM, 95–100.

[12] Ming Jin Cheok, Zaid Omar, and Mohamed Hisham Jaward. 2019. A review of hand gesture and sign language recognition techniques. *International Journal of Machine Learning and Cybernetics* 10, 1 (2019), 131–153.

[13] Mihaly Csikszentmihalyi and Keith Sawyer. 2014. *Shifting the Focus from Individual to Organizational Creativity.* Springer Netherlands, Dordrecht, 67–71. https://doi.org/10.1007/978-94-017-9085-7_6

[14] Mary Czerwinski, Eric Horvitz, and Edward Cutrell. 2001. Subjective duration assessment: An implicit probe for software usability. In *Proceedings of IHM-HCI 2001 conference*, Vol. 2. 167–170.

[15] Raimund Dachselt and Anett Hübner. 2007. Three-dimensional menus: A survey and taxonomy. *Computers & Graphics* 31, 1 (2007), 53–65.

[16] Kaushik Das and Christoph W Borst. 2010. An evaluation of menu properties and pointing techniques in a projection-based VR environment. In *2010 IEEE Symposium on 3D User Interfaces (3DUI).* IEEE, 47–50.

[17] Jing Du, Yangming Shi, Chao Mei, John Quarles, and Wei Yan. 2016. Communication by interaction: A multiplayer VR environment for building walkthroughs. In *Construction Research Congress 2016.* 2281–2290.

[18] Bruno Dumas, Denis Lalanne, and Sharon Oviatt. 2009. *Multimodal interfaces: A survey of principles, models and frameworks.* Springer, 3–26.

[19] Karin Eilers, Friedhelm Nachreiner, and Kerstin Hänecke. 1986. Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. *Zeitschrift für Arbeitswissenschaft* 4 (1986), 214–224.

[20] Martin Fischbach. 2017. *Enhancing Software Quality of Multimodal Interactive Systems.* Publication.

[21] Martin Fischbach, Dennis Wiebusch, and Marc Erich Latoschik. 2017. Semantic Entity-Component State Management Techniques to Enhance Software Quality for Multimodal VR-Systems. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 23, 4 (2017), 1342–1351. DOI: 10.1109/TVCG.2017.2657098.

[22] Sascha Gebhardt, Sebastian Pick, Franziska Leithold, Bernd Hentschel, and Torsten Kuhlen. 2013. Extended pie menus for immersive virtual environments. *IEEE transactions on visualization and computer graphics* 19, 4 (2013), 644–651.

[23] Jawaid A. Ghani and Satish P. Deshpande. 1994. Task characteristics and the experience of optimal flow in human—computer interaction. *The Journal of Psychology* 128, 4 (1994), 381–391.

[24] Beth A. Hennessey and Teresa M. Amabile. 2010. Creativity. *Annual Review of Psychology* 61, 1 (2010), 569–598. https://doi.org/10.1146/annurev.psych.093008.100416

[25] Arsalan Heydarian, Joao P Carneiro, David Gerber, Burcin Becerik-Gerber, Timothy Hayes, and Wendy Wood. 2015. Immersive virtual environments versus physical built environments: A benchmarking study for building design and user-built environment explorations. *Automation in Construction* 54 (2015), 116–126.

[26] Jörn Hurtienne, Katharina Weber, and Lucienne Blessing. 2008. Prior experience and intuitive use: image schemas in user centred design. In *Designing inclusive futures.* Springer, 107–116.

[27] Wijnand A IJsselsteijn, Yvonne AW de Kort, and Karolien Poels. 2013. The game experience questionnaire. *Eindhoven: Technische Universiteit Eindhoven* (2013), 3–9.

[28] IrisVR Inc. 2019. *Why This Public Utility Company Went from Unity & Unreal to Prospect for VR (And How Much It Saved Them).* Retrieved April 21, 2020 from https://blog.irisvr.com/navisworks-vr-case-study

[29] IrisVR Inc. 2020. *IrisVR.* Retrieved January 22, 2020 from https://irisvr.com/

[30] Seung-A A. Jin. 2011. "I feel present. Therefore, I experience flow:" A structural equation modeling approach to flow and presence in video games. *Journal of Broadcasting & Electronic Media* 55, 1 (2011), 114–136.

[31] Seung-A A. Jin. 2012. "Toward integrative models of flow": Effects of performance, skill, challenge, playfulness, and presence on flow in video games. *Journal of Broadcasting & Electronic Media* 56, 2 (2012), 169–186.

[32] Ed Kaiser, Alex Olwal, David McGee, Hrvoje Benko, Andrea Corradini, Xiaoguang Li, Phil Cohen, and Steven Feiner. 2003. Mutual Disambiguation of 3D Multimodal Interaction in Augmented and Virtual Reality. In *Proceedings of the 5th International Conference on Multimodal Interfaces.* Association for Computing Machinery, New York, NY, USA, 12–19. https://doi.org/10.1145/958432.958438

[33] James C Kaufman. 2012. Counting the muses: development of the Kaufman domains of creativity scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts* 6, 4 (2012), 298.

[34] James C. Kaufman, John Baer, Jason C. Cole, and Janel D. Sexton. 2008. A Comparison of Expert and Nonexpert Raters Using the Consensual Assessment Technique. *Creativity Research Journal* 20, 2 (2008), 171–178. https://doi.org/10.1080/10400410802059929

[35] James C. Kaufman and Vlad P. Glaveanu. 2019. *A Review of Creativity Theories: What Questions Are We Trying to Answer?* Cambridge University Press, 27–43. https://doi.org/10.1017/9781316979839.004

[36] Radwa Khalil, Ben Godde, and Ahmed A. Karim. 2019. The Link Between Creativity, Cognition, and Creative Drives and Underlying Neural Mechanisms. *Frontiers in Neural Circuits* 13 (2019), 18. https://doi.org/10.3389/fncir.2019.00018

[37] Kyung Hee Kim. 2006. Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity research journal* 18, 1 (2006), 3–14.

[38] Gregory Kramida. 2016. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE transactions on visualization and computer graphics* 22, 7 (2016), 1912–1931.

[39] Denis Lalanne, Laurence Nigay, philippe Palanque, Peter Robinson, Jean Vanderdonckt, and Jean-François Ladry. 2009. Fusion Engines for Multimodal Input: A Survey. In *Proceedings of the 2009 International Conference on Multimodal Interfaces.* Association for Computing Machinery, New York, NY, USA, 153–160. https://doi.org/10.1145/1647314.1647343

[40] Marc Erich Latoschik. 2001. A general framework for multimodal interaction in virtual reality systems: PrOSA. In *The Future of VR and AR Interfaces-Multimodal, Humanoid, Adaptive and Intelligent. Proceedings of the Workshop at IEEE Virtual Reality.* 21–25.

[41] Marc Erich Latoschik. 2002. Designing Transition Networks for Multimodal VR-Interactions Using a Markup Language. In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces.* IEEE Computer Society, USA, 411. https://doi.org/10.1109/ICMI.2002.1167030

[42] Marc Erich Latoschik. 2005. A user interface framework for multimodal VR interactions. In *Proceedings of the 7th international conference on Multimodal interfaces.* 76–83.

[43] Marc Erich Latoschik, Martin Fröhlich, Bernhard Jung, and Ipke Wachsmuth. 1998. Utilize Speech and Gestures to Realize Natural Interaction in a Virtual Environment. In *IECON'98: Proceedings of the 24th annual Conference of the IEEE Industrial Electronics Society*, Vol. 4. 2028–2033. http://trinity.inf.uni-bayreuth.de/download/usg_to_realize.pdf

[44] Marc Erich Latoschik and Christian Fröhlich. 2007. Semantic Reflection for Intelligent Virtual Environments. In *Proceedings of the IEEE VR 2007.* 305–306. https://downloads.hci.informatik.uni-wuerzburg.de/semantic-reflection-VR07.pdf

[45] Marc Erich Latoschik and Christian Fröhlich. 2007. Towards Intelligent VR: Multi-Layered Semantic Reflection for Intelligent Virtual Environments. In *Proceedings of the Graphics and Applications GRAPP 2007.* 249–259. https://downloads.hci.informatik.uni-wuerzburg.de/Towards-IVR-grapp07-latoschik.pdf

[46] Marc Erich Latoschik and Henrik Tramberend. 2010. Short Paper: Engineering Realtime Interactive Systems: Coupling & Cohesion of Architecture Mechanisms. In *Proceedings of the 16th Eurographics Conference on Virtual Environments & Second Joint Virtual Reality.* Eurographics Association, 25–28.

[47] M E. Latoschik and H Tramberend. 2011. Simulator X: A Scalable and Concurrent Architecture for Intelligent Realtime Interactive Systems. In *Proceedings of the 2011 IEEE Virtual Reality Conference.* IEEE Computer Society, USA, 171–174. https://doi.org/10.1109/VR.2011.5759457

[48] Joseph J. LaViola Jr, Ernst Kruijff, Ryan P. McMahan, Doug Bowman, and Ivan P. Poupyrev. 2017. *3D user interfaces: theory and practice.* Addison-Wesley Professional.

[49] Joseph J. LaViola Jr., Ernst Kruijff, Ryan P. McMahan, Doug A. Bowman, and Ivan P. Poupyrev. 2017. *3D User Interfaces – Theory and Practice* (2nd ed.). Addison-Wesley.

[50] Richard E Mayer. 1999. 22 Fifty years of creativity research. *Handbook of creativity* 449 (1999).
[51] Alexander S McKay, Maciej Karwowski, and James C Kaufman. 2017. Measuring the muses: validating the Kaufman domains of creativity scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts* 11, 2 (2017), 216.
[52] Microsoft. 2009. *Speech SDK*. Retrieved January 21, 2020 from https://www.microsoft.com/en-us/download/details.aspx?id=10121
[53] Jeanne Nakamura and Mihaly Csikszentmihalyi. 2014. The concept of flow. In *Flow and the foundations of positive psychology*. Springer, 239–263.
[54] Ali Bou Nassif, Ismail Shahin, Imtinan Attili, Mohammad Azzeh, and Khaled Shaalan. 2019. Speech recognition using deep neural networks: A systematic review. *IEEE Access* 7 (2019), 19143–19165.
[55] Anja Naumann and Jörn Hurtienne. 2010. Benchmarks for intuitive interaction with mobile devices. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. ACM, 401–402.
[56] Anja Naumann, Jörn Hurtienne, Johann Habakuk Israel, Carsten Mohs, Martin Christof Kindsmüller, Herbert A Meyer, and Steffi Hußlein. 2007. Intuitive use of user interfaces: defining a vague concept. In *International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer, 128–136.
[57] Laurence Nigay and Joëlle Coutaz. 1993. A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 172–178. https://doi.org/10.1145/169059.169143
[58] Sharon Oviatt. 1997. Multimodal Interactive Maps: Designing for Human Performance. *Hum.-Comput. Interact.* 12, 1 (March 1997), 93–129. https://doi.org/10.1207/s15327051hci1201&2_4
[59] Sharon Oviatt. 2003. Advances in robust multimodal interface design. *IEEE computer graphics and applications* 5 (2003), 62–68.
[60] Sharon Oviatt. 2006. Human-centered design meets cognitive load theory: designing interfaces that help people think. In *Proceedings of the 14th ACM international conference on Multimedia*. 871–880.
[61] Sharon Oviatt. 2012. Multimodal interfaces. In *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications, 3rd Edition*. Lawrence Erlbaum Assoc., Mahwah, NJ, 405–430.
[62] Sharon Oviatt and Philip Cohen. 2000. Perceptual User Interfaces: Multimodal Interfaces That Process What Comes Naturally. *Commun. ACM* 43, 3 (March 2000), 45–53. https://doi.org/10.1145/330534.330538
[63] Sharon Oviatt and Philip R Cohen. 2015. The paradigm shift to multimodality in contemporary computer interfaces. *Synthesis lectures on human-centered informatics* 8, 3 (2015), 1–243.
[64] Sharon Oviatt, Rachel Coulston, and Rebecca Lunsford. 2004. When Do We Interact Multimodally? Cognitive Load and Multimodal Communication Patterns. In *Proceedings of the 6th International Conference on Multimodal Interfaces*. Association for Computing Machinery, New York, NY, USA, 129–136. https://doi.org/10.1145/1027933.1027957
[65] Sharon Oviatt, Björn Schuller, Philip R. Cohen, Daniel Sonntag, Gerasimos Potamianos, and Antonio Krüger (Eds.). 2017. *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1*. Vol. 14. Association for Computing Machinery and Morgan & Claypool.
[66] Thies Pfeiffer and Marc Erich Latoschik. 2004. Resolving Object References in multimodal Dialogues for Immersive Virtual Environments. In *Proceedings of the IEEE Virtual Reality conference 2004*. 35–42. http://trinity.inf.uni-bayreuth.de/download/Resolving_Object_References.pdf
[67] Eeva M. Pilke. 2004. Flow experiences in information technology use. *International journal of human-computer studies* 61, 3 (2004), 347–357.
[68] Jonathan A. Plucker, Matthew C. Makel, and Meihua Qian. 2019. *Assessment of Creativity* (2 ed.). Cambridge University Press, 44–68. https://doi.org/10.1017/9781316979839.005
[69] Michael I. Posner. 1980. Orienting of attention. *Quarterly journal of experimental psychology* 32, 1 (1980), 3–25.
[70] Ronald Rosenfeld, Dan Olsen, and Alex Rudnicky. 2001. Universal speech interfaces. *interactions* 8, 6 (2001), 34–44.
[71] Mark A. Runco and Garrett J. Jaeger. 2012. The Standard Definition of Creativity. *Creativity Research Journal* 24, 1 (2012), 92–96. https://doi.org/10.1080/10400419.2012.650092 arXiv:https://doi.org/10.1080/10400419.2012.650092
[72] A Santos, Telmo Zarraonandia, Paloma Díaz, and Ignacio Aedo. 2017. A Comparative Study of Menus in Virtual Reality Environments. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. 294–299.
[73] Dimitrios Saredakis, Ancret Szpak, Brandon Birckhead, Hannah Keage, Rizzo Albert, and Tobias Loetscher. 2020. Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. *Frontiers in Human Neuroscience* 14 (2020). https://doi.org/10.3389/fnhum.2020.00096
[74] Prabir Sarkar, Amaresh Chakrabarti, et al. 2007. Development of a method for assessing design creativity. In *Proceedings of ICED 2007, the 16th International Conference on Engineering Design*. 349–350.
[75] Rajeev Sharma, Vladimir I Pavlović, and Thomas S Huang. 2002. Toward multimodal human–computer interface. In *Advances in image processing and understanding: a festschrift for Thomas S Huang*. World Scientific, 349–365.
[76] Nicholas Stefanic and Clint Randles. 2015. Examining the reliability of scores from the consensual assessment technique in the measurement of individual and small group creativity. *Music Education Research* 17, 3 (2015), 278–295. https://doi.org/10.1080/14613808.2014.909398
[77] Facebook Technologies. 2018. *Oculus Rift S*. Retrieved May 31, 2020 from https://www.oculus.com/rift-s/
[78] Unity Technologies. 2017. *Unity*. Retrieved January 18, 2020 from https://unity3d.com/
[79] Unity Technologies. 2020. *Unity*. Retrieved May 27, 2020 from https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@0.9/manual/index.html
[80] Unity Technologies. 2020. *Unity*. Retrieved May 27, 2020 from https://vrtoolkit.readme.io
[81] Ellis Paul Torrance. 1966. *Torrance tests of creative thinking: Norms-technical manual: Verbal tests, forms a and b: Figural tests, forms a and b*. Personal Press, Incorporated.
[82] John W Tukey. 1962. The future of data analysis. *The annals of mathematical statistics* 33, 1 (1962), 1–67.
[83] Karl T. Ulrich and Steven D. Eppinger. 2004. Product architecture. *Product design and development* 3 (2004), 163–186.
[84] VRTK. 2018. *Virtual Reality Toolkit*. Retrieved January 18, 2020 from https://www.vrtk.io/
[85] Erik Wolf, Sara Klüber, Chris Zimmerer, Jean-Luc Lugrin, and Marc Erich Latoschik. 2019. "Paint that object yellow": Multimodal Interaction to Enhance Creativity During Design Tasks in VR. In *2019 International Conference on Multimodal Interaction*. ACM, 195–204.
[86] Maliha Zaman, Murugan Anandarajan, and Qizhi Dai. 2010. Experiencing flow with instant messaging and its facilitating role on creative behaviors. *Computers in Human Behavior* 26, 5 (2010), 1009–1018.
[87] Ferdinand Rudolf Hendrikus Zijlstra. 1993. Efficiency in work behaviour: A design approach for modern tools. (1993).
[88] Chris Zimmerer, Martin Fischbach, and Marc Latoschik. 2018. Semantic Fusion for Natural Multimodal Interfaces using Concurrent Augmented Transition Networks. *Multimodal Technologies and Interaction* 2, 4 (2018), 81.
[89] Chris Zimmerer, Martin Fischbach, and Marc Erich Latoschik. 2018. *Concurrent Augmented Transition Network — Project Page*. Retrieved August 28, 2020 from https://www.hci.uni-wuerzburg.de/projects/mmi/
[90] Chris Zimmerer, Martin Fischbach, and Marc Erich Latoschik. 2018. Space Tentacles - Integrating Multimodal Input into a VR Adventure Game. In *Proceedings of the 25th IEEE Virtual Reality (VR) conference*. IEEE, 745–746. https://downloads.hci.informatik.uni-wuerzburg.de/2018-ieeevr-space-tentacle-preprint.pdf